Published in final edited form as:

Hum Mutat. 2019 July ; 40(7): 952–961. doi:10.1002/humu.23740.

Screening for rare epigenetic variations in autism and schizophrenia

Paras Garg¹, Andrew J. Sharp¹

Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, Hess Center for Science and Medicine, New York, New York

Abstract

While many studies have led to the identification of rare sequence variants linked with susceptibility to autism and schizophrenia, the contribution of rare epigenetic variations (epivariations) in these disorders remains largely unexplored. Previously we presented evidence that epivariations occur relatively frequently in the human genome, and likely contribute to a subset of congenital and neurodevelopmental disorders through the disruption of dosage sensitive genes. Here we extend this approach, studying methylation profiles from 297 samples with autism and 767 cases with schizophrenia, identifying 84 and 268 rare epivariations in these two cohorts, respectively, that were absent from 4,860 population controls. We observed multiple features associated with these epivariations that support their pathogenic relevance, including (i) a significant enrichment for epivariations in schizophrenic individuals at genes previously linked with schizophrenia, (ii) increased brain expression of genes associated with epivariations found in autism cases compared to controls, (iii) in autism families, a significant excess of epivariations found specifically in affected versus unaffected sibs, (iv) Gene Ontology terms linked with epivariations found in autism including “D1 dopamine receptor binding”. Our study provides additional evidence that rare epivariations likely contribute to the mutational spectra underlying neurodevelopmental disorders.

Keywords

Epigenetics; Epimutation; DNA methylation; Epivariation; Tandem repeat expansion

INTRODUCTION

Autism and schizophrenia are common neurological disorders that are estimated to affect 1% and 0.75% of the worldwide population, respectively (Bassett & Costain, 2012). While both disorders are highly polygenic, microarray and exome sequencing studies have enabled the discovery of multiple highly-penetrant genetic risk factors in these disorders, including rare inherited and *de novo* single nucleotide variants (SNVs) (De Rubeis et al., 2014; Genovese et al., 2016; McCarthy et al., 2014; Mowry & Gratten, 2013; O’Dushlaine et al.,

¹Address for correspondence: Andrew J. Sharp, Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, Hess Center for Science and Medicine, 1470 Madison Avenue, Room 8-116, Box 1498, New York, NY 10029 USA. Telephone: +1-212-824-8942, Fax: +1-646-537-8527, andrew.sharp@mssm.edu.

2014; O’Roak et al., 2011; Sklar et al., 2014; Wormley et al., 2016), and copy number variations (De Rubeis & Buxbaum, 2015; Szatkiewicz et al., 2014). For example, the Autism Sequencing Consortium (ASC) identified rare *de novo* mutations that caused loss of function in 107 genes for ~5% of autism cases (De Rubeis et al., 2014). Similarly, studies in schizophrenia have shown an enrichment of rare coding variants in cases compared to controls, leading to a polygenic burden that increases risk (Hannon et al., 2018; O’Dushlaine et al., 2014). Furthermore, there is considerable evidence for a shared genetic etiology of schizophrenia, autism and intellectual disability (Kushima et al., 2018; Li et al., 2016; McCarthy et al., 2014; Shohat et al., 2017).

While the focus of genetic studies of human disease has been on sequence variation, the contribution of epigenetic causes remains an area that is largely unexplored. However, recent studies have now demonstrated that in some human disorders, epigenetic changes represent the pathogenic defect in diseases that have traditionally been regarded as purely genetic in nature. For example, promoter hypermethylation of the *BRCA1* gene (MIM# 113705) leads to transcriptional silencing of the gene in some patients with hereditary breast and ovarian cancer (Evans et al., 2018). Similarly, promoter methylation of *MLH1* (MIM# 120436) has been found in ~5% of pedigrees with hereditary colon cancer (Castillejo et al., 2015), and compound heterozygosity for a coding mutation and promoter methylation of the *MMACHC* (MIM# 609831) gene has recently been reported in patients with autosomal recessive inborn errors of vitamin B₁₂ metabolism (Guéant et al., 2018). These observations suggest that epigenetic defects may be a general mutational mechanism that contributes more broadly to the mutational spectra of many human disorders.

Recently we reported findings in a cohort of patients with idiopathic congenital disorders, intellectual disability and/or autism, where we identified rare epigenetic variations in 23% of these cases. Significantly, *de novo* epivariations occurred at an increased rate in these cases compared to controls, and were often associated with expression defects of the associated genes (Barbosa et al., 2018). Here we extend this approach to study >1,000 individuals with autism and schizophrenia, identifying hundreds of rare epivariations in these cases, and providing further evidence that these epigenetic defects likely underlie some patients with neurodevelopmental disorders.

MATERIALS AND METHODS

We obtained DNA methylation data generated using the Illumina 450k HumanMethylation BeadChip (450k array) from three published cohorts, all of which comprised DNA methylation profiles from whole blood. Cohort 1 [GSE80417] comprised 353 schizophrenia cases and 322 controls, while cohort 2 [GSE84727] comprised 414 schizophrenia cases and 433 controls (Hannon et al., 2016), both of which were downloaded from GEO (<http://www.ncbi.nlm.nih.gov/geo/>). We also utilized methylation profiles of 364 autism cases and 364 unaffected siblings, downloaded from dbGaP study phs000619.v1.p1. These samples were compared against a control cohort of 1,534 unrelated individuals from the general population, which represents the merger of four separate datasets that had similarly undergone profiling of peripheral blood DNA with the Illumina 450k array [GSE36064, GSE40279, GSE42861, GSE53045]. We further incorporated data from an additional 2,711

population controls [GSE55763], and 117 families from GSE56105 (Shah et al., 2014), as described in (Barbosa et al., 2018).

Data filtering and normalization

Each data set underwent several filtering and normalization steps, as described in (Barbosa et al., 2018). In short, 482,421 probe sequences (50-mer oligonucleotides) were remapped to the reference human genome hg19 (NCBI37) using BSMAP, allowing up to 2 mismatches and 3 gaps, and we retained only uniquely mapping autosomal probes. We removed any probe that overlapped SNVs with MAF $\geq 5\%$ identified by the 1000 Genomes Project within 5 bp upstream or including the targeted CpG. Further, we removed probes with detection $p > 0.01$ (an indicator of low signal intensity) in each individual. We generated PCA plots using raw β -values for autosomal probes, and removed outlier individuals based on PC1 and PC2. After filtering, we retained 297 probands and 286 siblings from the autism cohort, and 767 cases and 754 controls from the schizophrenia cohorts. We applied background correction, two color channel normalization and quantile normalization using the *lumi* package in R (Du, Kibbe, & Lin, 2008). The distribution of Infinium I and Infinium II probes were adjusted using BMIQ (Teschendorff et al., 2013). Probes were then annotated based on their position relative to RefSeq genes using BEDTools v2.17. 1,521 schizophrenia cases and controls were normalized together with the 1,534 controls, while for the autism cohort, we normalized cases and siblings together as a group, and then quantile normalized these β -values with those of the 1,534 controls.

Epivariation analysis

We utilized a sliding window approach to identify epivariations in each sample, as described in (Barbosa et al., 2018). Methylation profiles of each sample were compared against the cohort of 1,534 controls, detecting regions of outlier methylation absent in the controls represented by clusters of multiple independent probes with extreme methylation values. Stringent thresholds for calling Differentially Methylated Regions (DMRs) were set as follows: Hypermethylated DMRs where a case presented, in a 1kb window, at least 3 probes that each had β -values above the 99.9th percentile of the control distribution for that probe and are ≥ 0.15 above the control mean, and at least one of these probes had a β -value ≥ 0.1 above the maximum observed in controls for that probe. Similarly, hypomethylated DMRs where a case presented, in a 1kb window, at least three probes that each had β -values below the 0.1th percentile of the control distribution for that probe, and are ≥ 0.15 below the control mean, and at least one of these probes had a β -value ≤ 0.1 below the minimum observed in controls for that probe. 767 schizophrenia cases were compared against methylation profiles of 754 controls combined with the 1,534 external controls, while 297 autism cases and their 286 siblings were compared against methylation profiles of 1,534 external controls.

All DMRs were manually curated to remove loci that were deemed false-positive calls as described in (Barbosa et al., 2018). Despite performing probe level filtering and multiple rounds of normalization of the array data, such measures are imperfect and do not remove all probes that show aberrant signals due to underlying technical or biological effects. We observed both systematic batch effects, and also sporadic false positives in single samples that were filtered, as follows:

- i. Batch effects, i.e., technical differences due to arrays being processed in separate groups, were sometimes observed between cases and controls. Here, it was usually observed that there was either a systematic shift in β -values reported by one or more probes within a region between arrays processed in different batches. In some cases the mean of each batch was significantly different, with every sample showing a shift, whereas in other cases the means of the two populations remained similar, but a subset of the samples in one batch showed a gradient of deviations, with the β -values of multiple cases lying in the extreme tail of the control distribution.
- ii. In some cases while 3 probes within a 1-kb region were identified as outliers, the outlier probes were not in a contiguous block as would be expected for a true methylation change, and were interspersed with other probes that showed no difference compared to the control population. We interpreted these signals as likely random groupings of individual probes that each yielded outlier beta values for some other reason, e.g., rare underlying variations that influenced probe performance, or poor hybridization performance of individual probes.

In our previous study, after automated identification according to the above criteria followed by manual curation, subsequent validation studies indicated a 95% true positive rate for differentially methylated regions (DMRs) detected using this pipeline, thus showing that we are highly specific for identifying regions of true outlier methylation.

We removed four samples from the schizophrenia cohort that each presented with >10 DMRs, as in our previous analysis we found that such samples were typically low quality. In order to identify DMRs that were specific to samples with autism or schizophrenia, we then overlapped DMRs with those that were observed two other cohorts of population controls (Barbosa et al., 2018).

As regions of homozygous deletion can cause outlier methylation values (Barbosa et al., 2018), we performed a sliding window analysis to detect clusters of probes with failed detection p-values in each sample. We used a 1 kb sliding window analysis to identify regions with 3 probes with detection p-value >0.01.

Annotations and enrichment analysis

For autism we utilized two gene sets: (i) 100 genes showing a significant enrichment (FDR 10%) of *de novo* damaging missense and nonsense mutations in autism probands by the Autism Sequencing Consortium (Joseph Buxbaum, personal communication), and (ii) 1,019 genes implicated in autism from a variety of sources (downloaded from <https://www.sfari.org/resource/sfari-gene/>). For schizophrenia we utilized two gene sets: (i) 2,268 genes implicated in schizophrenia risk from exome sequencing and CNV studies in schizophrenia, GWAS studies of schizophrenia, and genes contained within relevant Gene Ontology categories and pathways (O'Dushlaine et al., 2014), and (ii) a more stringent subset of this list comprising 930 genes that have been observed as having *de novo* LOF variants, *de novo* non-synonymous variants, or occurred within *de novo* CNVs in schizophrenia cases. For enrichment analyses of autism and schizophrenia genes, we utilized a background set comprising 38,643 collapsed autosomal regions representing all 1kb

windows on the 450k array that contained 3 probes and were effectively assayed for epivariations. DMRs were associated to Refseq genes based on overlap with gene body and/or ± 2 kb of their transcription start site. Enrichment analyses were performed by comparing against the set of all autosomal probes that lie in the background set, and p-values generated using the Hypergeometric distribution (phyper function in R). Fold enrichments were calculated using the following formula: (probes in DMRs overlapping Feature/probes in DMRs)/(probes in Background overlapping Feature/probes in Background).

To assess brain expression, the set of genes associated with DMRs were annotated with their mean gene expression level in 12 different brain regions generated by the GTEx project based on overlap with gene body and/or ± 2 kb of their transcription start site (Berghuis et al., 2013). Fold changes in gene expression were calculated by comparison of median expression of genes associated with epivariations against the median expression of genes associated with the set of all autosomal probes that lie in the background set on the array, as outlined above. Permutation p-values for expression change in the autism and schizophrenia cohorts were generated by taking the set of DMRs defined in that disease cohort together with the set of all DMRs defined in the 2,711 controls, randomly selecting the same number of DMRs as found in the disease cohort from this total set, and the p-value calculated from the number of 10,000 random permutations that showed mean expression greater than or equal to that observed in the disease cohort.

To identify enriched functional categories related to genes associated with DMRs in autism and schizophrenia, we performed Gene Ontology (GO) analysis using the method described in Eden *et al.* (Yakhini et al., 2009).

Results

Utilizing a robust sliding window approach followed by manual curation, we identified a total of 105 DMRs in 76 autism cases (25.6% of all autism cases) (Aut-DMRs) and 309 DMRs in 220 schizophrenia cases (28.6% of all schizophrenia cases) (Scz-DMRs) (Supp. Tables S1 and S2). Of these, five Aut-DMRs and three Scz-DMRs overlapped with clusters of probes with failed detection p-values, suggesting that these are not true epigenetic defects, but instead likely result from the presence of underlying homozygous deletions (Supp. Figure S1, Supp. Table S3). Overlap of Aut-DMRs and Scz-DMRs with epivariations identified in 3,326 population controls from two cohorts we studied previously (Barbosa et al., 2018), showed that 16 of the Aut-DMRs and 38 of the Scz-DMRs were also present in the general population. After removing these, we utilized a final set of 84 Aut-DMRs in 63 autism cases and 268 Scz-DMRs in 200 schizophrenia cases for further analysis (Supp. Tables S1 and S2).

After discarding DMRs that were also seen in controls, or were due to underlying homozygous deletions, there was a small but non-significant burden of rare DMRs in autism cases compared to their unaffected sibs (91 DMRs in 297 cases versus 77 DMRs in 286 unaffected sibs, 1.13-fold enrichment, $p=0.36$). However, as expected, many of these DMRs were shared between both affected and unaffected sibs. When we considered only those DMRs that were found specifically in autism cases but absent in their unaffected sibs, and

compared these to the number of DMRs that were found only in their unaffected sibs but were absent in cases, there was a significant excess of DMRs found specifically in affected versus unaffected sibs (59 DMRs unique to 297 cases versus 35 DMRs unique to 286 unaffected sibs, 1.62-fold enrichment, $p=0.013$, two-tailed Fisher's Exact test).

Clusters of rare epivariations in samples with autism and schizophrenia

We observed two cases of autism and four cases of schizophrenia in whom multiple independent epivariations occurred in a single individual that were separated by <200 kb (Figure 1, Supp. Tables S1 and S2). In some instances, these clustered epivariations included both gains and losses of methylation at different loci, suggesting a general disruption of epigenetic state over entire domains. The frequency with which these clustered epivariations occurred was 2.1-fold higher in samples with autism/schizophrenia compared to population controls, although this did not reach statistical significance (6 of 1,064 cases versus 9 of 3,326 controls, $p=0.22$, two-tailed Fisher's Exact test).

In several cases these clustered epivariations occurred at genes with known links to neurodevelopmental disorders, further suggesting that these events may contribute to patient phenotype. For example, in a sample with autism we observed hypermethylation of a CpG island immediately upstream of *GIGYF1* (MIM# 612064), and a ~1.5kb region of hypomethylation within the gene body (chr7:100,281,505–100,281,760 and chr7:100,290,844–100,292,330). Both of these epivariations were absent in the patient's unaffected sibling, and *GIGYF1* has been identified by the ASC as showing a significant enrichment for damaging coding variants in samples with autism (J. Buxbaum, personal communication). We also observed one individual with autism who had two regions of hypomethylation intragenic within *SHANK3* (MIM# 606230)(chr22:51,158,549–51,159,556 and chr22:51,169,027–51,170,004), a gene which shows recurrent mutations in patients with autism (Boccuto et al., 2013). While these epivariations in *SHANK3* were absent in the unaffected sib of this autism case, they were identified in one control sample (out of the total of 4,860 population controls utilized overall), and thus we did not include this DMR in downstream enrichment analyses.

Similarly in samples with schizophrenia, we observed three CpG islands in a 60kb interval on 8q24.3, located upstream of *LYNX1* (MIM# 606110), *THEM6* and *LOC100288181*, that were all hypermethylated in an individual with schizophrenia. *LYNX1* is predominantly expressed in the brain, shares characteristics with toxins that bind and inhibit nicotinic acetylcholine receptors, and has been implicated in maintaining the stability of cortical networks and neuronal plasticity (Miwa et al., 2006; Morishita, Miwa, Heintz, & Hensch, 2010).

An enrichment of rare epivariations at imprinted loci

Consistent with previous studies (Aref-Eshghi et al., 2017; Joshi et al., 2016), we observed multiple epigenetic defects at imprinted loci in cases with autism and schizophrenia. The fraction of epivariations occurring at imprinted loci were significantly enriched in both autism and schizophrenia cases when compared to the overall genomic distribution of epivariations in these two disorders: five of 84 DMRs observed in autism were at imprinted

loci (*INPP5F* (MIM# 609389), *NDN* (MIM# 602117), *GNAS* (MIM# 139320), *HYMAI/PLAGL1* (MIM# 606546/603044) and *MEST* (MIM# 601029)) (62-fold enrichment versus background, $p=1.7\times 10^{-6}$), and six of 268 DMRs observed in schizophrenia were at imprinted loci (*MEG3* (MIM# 605636), *TUBGCP5* (MIM# 608147), *MAGEL2* (MIM# 605283), *NDN*, *HM13/MCTS2P* (MIM# 607106), and *PEG10* (MIM# 609810)) (23-fold enrichment versus background, $p=2\times 10^{-7}$) (Supp. Figure S2). Specific imprinting anomalies included the following: (i) One individual with schizophrenia showed gain of methylation of the *PEG10* imprinted locus. (ii) One individual with schizophrenia showed hypomethylation at both *MAGEL2* and *NDN*, both of which lie within the Prader-Willi/Angelman syndrome imprinted region at 15q11.2, and are imprinted with parental-specific methylation. (iii) One individual with autism also showed hypomethylation of *NDN* in the 15q11.2 imprinted region.

An enrichment of rare epivariations at genes linked to autism and schizophrenia

We tested whether epivariations identified in samples with autism and schizophrenia were enriched for genes with known links to these conditions from prior genetic studies. For autism we utilized two gene sets: (i) autism gene set A, containing 100 genes with significant enrichments for *de novo* damaging coding SNVs in autism cases identified by exome sequencing studies (J. Buxbaum, personal communication), and (ii) autism gene set B, containing 1,019 genes implicated in autism from a variety of sources (downloaded from <https://www.sfari.org/resource/sfari-gene/>). For schizophrenia we utilized two gene sets: (i) schizophrenia gene set A, containing 2,268 genes with putative links to schizophrenia from a variety of sources, including exome sequencing and copy number variation (CNV) studies in schizophrenia, GWAS studies of schizophrenia, and genes contained within relevant Gene Ontology categories and pathways (O'Dushlaine et al., 2014), and (ii) schizophrenia gene set B, representing a more stringent subset of this list comprising 930 genes that have been observed as having *de novo* LOF variants, *de novo* non-synonymous variants, or occurred within *de novo* CNVs in schizophrenia cases.

In autism gene set A we identified two epivariations that were associated with the set of 100 causative autism genes, representing a 4.2-fold enrichment compared to background, although this does not achieve statistical significance ($p=0.08$). These two genes were *SKI* (MIM #164780) and *GIGYF1* (MIM #612064). Autism gene set B overlapped 5 epivariations, representing a 1.18-fold enrichment compared to background, which was also not significant ($p=0.42$). In schizophrenia gene set A, we identified epivariations at 52 genes with putative links to schizophrenia, representing a 1.58-fold enrichment versus background ($p=9.4\times 10^{-4}$). Using the more stringent schizophrenia gene set B, we observed a 1.89-fold enrichment for DMRs found in schizophrenia cases to overlap these genes ($p=0.0015$).

Genes associated with rare epivariations in autism cases show increased expression in the brain

We hypothesized that genes that influence neurodevelopmental disorders might be enriched for brain expression compared to the null. Furthermore, as we utilized epigenome profiles obtained from DNA derived from whole blood to study neurodevelopmental disorders, this raises the question as to whether the epigenetic changes we detected in blood are relevant to

autism and schizophrenia. Therefore, to assess the potential pathogenic relevance of the epivariations we identified in cases of autism and schizophrenia, we utilized gene expression data from the GTEx project (<https://www.gtexportal.org/home/>). The set of genes associated with epivariations found in blood were annotated with their mean gene expression level in 12 different brain regions generated by the GTEx project. We then compared the expression level of the genes associated with epivariations identified in autism samples, schizophrenia samples, and 2,711 population controls versus the background set of all genes effectively tested on the array. While epivariations found in population controls showed no enrichments for high brain expression (mean fold-change versus background over all brain regions=1.02), we observed significantly higher expression in eight brain regions for genes associated with epivariations in autism cases (mean fold-change versus background over all brain regions=1.56 in autism) (Figure 3, Supp. Table S4). However, for genes associated with rare epivariations in schizophrenia cases we observed no significant changes versus background.

Gene Ontology terms associated with rare epivariations in autism

To gain additional insight into the function of genes associated with epivariations, we performed Gene Ontology (GO) analysis. The two most significant terms for Aut-DMRs were “Genetic imprinting” and “D1 dopamine receptor binding” with 34-fold (raw $p=8.7\times10^{-5}$, FDR $q=0.21$) and 57-fold ($p=5.2\times10^{-4}$, FDR $q=0.24$) enrichments, respectively (Supp. Figure S3, Supp. Table S5). For Scz-DMRs no GO terms showed FDR $q<0.5$ (Supp. Table S6).

DISCUSSION

Here we have performed a survey for rare epigenetic variation in >1,000 samples with a diagnosis of either autism or schizophrenia, identifying hundreds of regions of outlier methylation that were not observed in >4,800 population controls. In contrast to epigenome-wide association studies (EWAS) that compare populations of cases and controls, and which will identify sites that exhibit consistent but subtle differences between entire cohorts, our analysis approach looks for large methylation changes found in single individuals, and therefore identifies rare events that are missed by EWAS.

Similar to our previous study in a cohort of samples with a variety of congenital disorders, intellectual disability and/or autism, multiple lines of evidence suggest that these epivariations may play a role in the etiology of human disease. In particular, we observed (i) significant enrichments for epivariations at genes known to be linked with autism and schizophrenia in prior genetic studies, (ii) an enrichment for genes with high brain expression associated with these rare epivariations in autism, (iii) an increased incidence of epivariations specific to affected versus unaffected sibs in families with autism, (iv) that one of the most significant Gene Ontology term associated with rare epivariations in autism was “D1 dopamine receptor binding”. We also identified several examples of clustered epivariations, where multiple distinct regions of epigenetic change separated by 10–200kb occurred in a single individual. Notably, many of these clustered epivariations occurred at genes with known or potential roles in neurodevelopmental phenotypes. One plausible hypothesis is that these clustered epivariations correspond to more severe genomic

dysregulation of local regions that is more likely to result in disease. However, it should be noted that the overall fraction of cases of autism and schizophrenia that can potentially be explained by epivariations remains relatively small: no epivariations were identified in the majority of cases tested, and only a subset of epivariations occur at genes with known links to neurodevelopment. Also we note that while we observed an enrichment for genes with high brain expression associated with rare epivariations found in autism cases, this was not the case for epivariations identified in schizophrenia samples. The reasons underlying this are unclear.

Consistent with several previous studies (Aref-Eshghi et al., 2017; Barbosa et al., 2018; Kolarova et al., 2015), we observed that epigenetic changes occurred with increased frequency at imprinted loci. One sample with a diagnosis of autism showed hypomethylation of *NDN*, while a second individual with schizophrenia showed hypomethylation at both *NDN* and the neighboring imprinted gene *MAGEL2*. Aberrant imprinting of *NDN* has previously been linked with disruption of serotonergic neurons in mice (Rieusset et al., 2013), and thus it is possible that these imprinting anomalies contribute to the diagnosis of autism and schizophrenia. It should be noted however, that due to the allele-specific nature of methylation at imprinted loci, it is possible that the presence of an underlying heterozygous deletion at these regions could be incorrectly interpreted as an epivariation.

Many of the other imprinted genes where we observed methylation defects are not normally associated with neurocognitive deficits. Indeed, several autism samples and their siblings showed loss of methylation across multiple imprinted loci. This observation suggests a potential underlying cause in these families, and indeed multilocus imprinting disturbance of offspring is caused by mutations in several maternal effect proteins (Sanchez-Delgado et al., 2016; Turner et al., 2018).

We also identified several instances of DMRs that were apparently driven by clusters of probes with failed detection p-values, which likely indicate an underlying homozygous deletion at that region. For example, we detected a ~10kb DMR within the *SNORD* gene cluster at 15q11.2 in a sample with autism, where probes comprising the DMR corresponded perfectly to clusters of probes with failed detection p-values (Supp. Figure S1). Although we did not consider these signals as true epigenetic variations, these observations do allow the indirect identification of regions of homozygous deletion, which in some cases might be relevant to the phenotype of the individual. For example, deletion of the *SNORD* gene cluster has been implicated in the etiology of Prader-Willi syndrome (Person et al., 2010), and therefore this deletion could conceivably be related to a diagnosis of autism.

Several different factors likely underlie the generation of epivariations in the human genome. While some epivariations apparently occur sporadically, others are known to be secondary events that are caused by an underlying genetic aberration (Horsthemke, 2006). Examples of this latter type include mutation of nearby regulatory motifs, e.g. CTCF binding sites (Barbosa et al., 2018), CNVs that modify local regulatory elements (Cini et al., 2015; Barbosa et al., 2018), or expansions of GC-rich tandem repeats (Hansen et al., 1992; Winnepenninckx et al., 2007; Liu et al., 2014). Although we did not have access to DNA from the samples we surveyed, to investigate the possibility that tandem repeat expansions

might underlie some of the epivariations we identified, we performed an overlap of these with annotated tandem repeats in the genome (Simple Repeats track, UCSC Genome Browser). We observed that four of the epivariations identified in autism cases overlapped tandem repeats with a CGG motif (*NOG* (MIM# 602991), *GIPC1* (MIM# 605072), *GIGYF1* and *FRA10AC1* (MIM# 608866)) (Supp. Table S1). All four were hypermethylation events, and of these, hypermethylation of *FRA10AC1* is known to be caused by an underlying CGG expansion (Sarafidou et al., 2004). Similarly, three of the epivariations identified in schizophrenia samples overlapped CGG repeats (*NIPA1* (MIM# 608145), *ZNF713* (MIM# 616181), *KLF14* (MIM# 609393)), and in addition four others overlapped repeats with CCCC, CGCC or CCCCC motifs (*ECHDC3*, *B4GALNT4*, *MAP3K5* (MIM# 602448), *CCM2L*) (Supp. Table S2). Similarly, all of these epivariations were also hypermethylation events. Interestingly, hypermethylated expansions of the CGG repeat at the 5' end of *ZNF713* have previously been reported in association with autism in two families (Metsu et al., 2014). Although we cannot be sure that the epivariation we observe in this schizophrenia patient is caused by the same CGG expansion, it seems plausible that this is the case, therefore suggesting that expansions of this repeat likely predispose to both autism and schizophrenia. These observations indicate that at least some of the epivariations we detect are likely due to underlying expansions of GC-rich tandem repeats, and suggest several other novel epivariation loci that may result from underlying repeat expansions, warranting further studies of these loci.

One of the major limitations of our study is that we do not have access to biological samples from any of the individuals in this study, and are therefore unable to perform experimental verification of the epivariations detected by array. While it is possible that some of the loci we report might be false positives due to eg. hybridization artifacts, in our previous study which used an identical pipeline to identify epivariations, by performing bisulfite sequencing validation assays we showed a 95% true positive rate for DMRs we identified from the 450k array using our outlier approach, indicating that the vast majority of our calls are likely robust (Barbosa et al., 2018). Similarly, we do not have access to parental data, and are therefore unable to determine the inheritance patterns of epivariations. In addition, we chose to define epivariations using a sliding window approach that requires a cluster of at least three outlier probes, as this vastly increases robustness for identifying genuine epigenetic outliers. For example, single outlier probes that are caused by a variety of underlying biological or technical effects occur at relatively high frequency, and would result in a high false positive rate. However, the drawback of this approach is that we only sampled regions on the array that contain at least 3 probes clustered within a region of 1kb, limiting our assessment of epivariations to those loci that contain multiple closely-spaced probes. As a result, our methodology ignores data from ~31% of probes on the array that do not lie in such clusters, preferentially sampling CpG islands and shores, and under-sampling CpG shelves and intergenic regions. It is therefore likely that due to the limited coverage of the Illumina 450k array and our analysis approach, there are additional regions of epigenetic change that are missed in our analysis. The use of the methods with improved coverage, such as the newer 850k array, or whole-genome approaches such as bisulfite sequencing would provide improved power, although to our knowledge, no such data from sufficiently large populations exists at this time.

In addition, the methylation profiles utilized here were obtained from peripheral blood DNA, whereas both autism and schizophrenia are disorders of the brain. However, our previous studies (Barbosa et al., 2018) showed that epivariations detected in blood are generally conserved across multiple different tissues within an individual. Thus, although the study of methylation in blood derived DNA is not optimal in the case of neurodevelopmental disorders, the available evidence suggests that the majority of epivariations are constitutive events, suggesting that the use of blood is likely to be a valid approach.

In summary, our analysis of methylomes from >1,000 individuals with a diagnosis of autism or schizophrenia provides strong support for the hypothesis that rare epigenetic lesions contribute to the etiology of neurodevelopmental disorders. Our observations are consistent with a model in which epivariations represent a general mechanism of genomic dysregulation, raising the notion that they may contribute to many different human phenotypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported by NIH grant HG006696 to AJS. Research reported in this paper was supported by the Office of Research Infrastructure of the National Institutes of Health under award number S10OD018522. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was supported in part through the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

The authors would like to thank R. Alisch, P. Chopra, B. Barwick, and S. Warren, as well as a Simons Foundation (SFARI) award and NIH grant MH089606, both to S. Warren. We are grateful to all the families at the participating SFARI Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, E. Wijsman).

Grant numbers (as applicable-to ensure proper identification of funders with publication requirements-see note under Author Licensing; below): This work was supported by NIH grant HG006696 to AJS.

REFERENCES

- Aref-Eshghi E, Schenkel LC, Lin H, Skinner C, Ainsworth P, Paré G, ... Sadikovic B (2017). Clinical Validation of a Genome-Wide DNA Methylation Assay for Molecular Diagnosis of Imprinting Disorders. *Journal of Molecular Diagnostics*, 19, 848–856. [PubMed: 28807811]
- Barbosa M, Joshi RS, Garg P, Martin-Trujillo A, Patel N, Jadhav B, ... Sharp AJ (2018). Identification of rare de novo epigenetic variations in congenital disorders. *Nature Communications*, 9, 2064.
- Bassett A, & Costain. (2012). Clinical applications of schizophrenia genetics: genetic diagnosis, risk, and counseling in the molecular era. *The Application of Clinical Genetics*, 5, 1. [PubMed: 23144566]
- Berghuis B, Shad S, Ardlie K, Wen X, Liu J, Jewell S, ... Salvatore M (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45, 580–585. [PubMed: 23715323]
- Boccuto L, Lauri M, Sarasua SM, Skinner CD, Buccella D, Dwivedi A, ... Schwartz CE (2013). Prevalence of SHANK3 variants in patients with different subtypes of autism spectrum disorders. *European Journal of Human Genetics*, 21, 310–316. [PubMed: 22892527]
- Castillejo A, Egoavil C, Llor X, Andreu M, Payá A, Hernández-Illán E, ... Jover R (2015). Prevalence of MLH1 constitutional epimutations as a cause of Lynch syndrome in unselected versus selected

- consecutive series of patients with colorectal cancer. *Journal of Medical Genetics*, 52, 498–502. [PubMed: 25908759]
- Cini G, Carnevali I, Quaia M, Chiaravalli AM, Sala P, Giacomini E, Maestro R, Tibiletti MG, Viel A (2015). Concomitant mutation and epimutation of the MLH1 gene in a Lynch syndrome family. *Carcinogenesis*, 36, 452–8. [PubMed: 25742745]
- De Rubeis S, & Buxbaum JD (2015). Genetics and genomics of autism spectrum disorder: embracing complexity. *Human Molecular Genetics*, 24, R24–R31. [PubMed: 26188008]
- De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, ... Buxbaum JD (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, 515, 209–215. [PubMed: 25363760]
- Du P, Kibbe WA, & Lin SM (2008). lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, 24, 1547–1548. [PubMed: 18467348]
- Evans DGR, van Veen EM, Byers HJ, Wallace AJ, Ellingford JM, Beaman G, ... Newman WG (2018). A Dominantly Inherited 5' UTR Variant Causing Methylation-Associated Silencing of BRCA1 as a Cause of Breast and Ovarian Cancer. *The American Journal of Human Genetics*, 103, 213–220. [PubMed: 30075112]
- Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M, ... McCarroll SA (2016). Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nature Neuroscience*, 19, 1433–1441. [PubMed: 27694994]
- Guéant J-L, Chéry C, Oussalah A, Nadaf J, Coelho D, Josse T, ... Rosenblatt DS (2018). A PRDX1 mutant allele causes a MMACHC secondary epimutation in cblC patients. *Nature Communications*, 9, 67.
- Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, ... Mill J (2016). An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biology*, 17, 176. [PubMed: 27572077]
- Hannon E, Schendel D, Ladd-Acosta C, Grove J, Hansen CS, Andrews SV, ... Mill J (2018). Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Medicine*, 10, 19. [PubMed: 29587883]
- Hansen RS, Gartler SM, Scott CR, Chen SH, Laird CD (1992). Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. *Human Molecular Genetics*, 1, 571–8. [PubMed: 1301165]
- Horsthemke B (2006). Epimutations in human disease. *Current Topics in Microbiology and Immunology*, 310, 45–59. [PubMed: 16909906]
- Joshi RS, Garg P, Zaitlen N, Lappalainen T, Watson CT, Azam N, ... Sharp AJ (2016). DNA Methylation Profiling of Uniparental Disomy Subjects Provides a Map of Parental Epigenetic Bias in the Human Genome. *The American Journal of Human Genetics*, 99, 555–566. [PubMed: 27569549]
- Kolarova J, Tangen I, Bens S, Gillessen-Kaesbach G, Gutwein J, Kautza M, ... Caliebe A (2015). Array-based DNA methylation analysis in individuals with developmental delay/intellectual disability and normal molecular karyotype. *European Journal of Medical Genetics*, 58, 419–425. [PubMed: 26003415]
- Kushima I, Aleksic B, Nakatochi M, Shimamura T, Okada T, Uno Y, ... Ozaki N (2018). Comparative Analyses of Copy-Number Variation in Autism Spectrum Disorder and Schizophrenia Reveal Etiological Overlap and Biological Insights. *Cell Reports*, 24, 2838–2856. [PubMed: 30208311]
- Li J, Cai T, Jiang Y, Chen H, He X, Chen C, ... Wu J (2016). Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Molecular Psychiatry*, 21, 290–297. [PubMed: 25849321]
- Liu EY, Russ J, Wu K, Neal D, Suh E, McNally AG, Irwin DJ, Van Deerlin VM, Lee EB (2014). C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. *Acta Neuropathology*, 128, 525–41.
- McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y, ... Corvin A (2014). De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Molecular Psychiatry*, 19, 652–658. [PubMed: 24776741]

- Metsu S, Rainger JK, Debacker K, Bernhard B, Rooms L, Grafodatskaya D, Weksberg R, Fombonne E, Taylor MS, Scherer SW, Kooy RF, FitzPatrick DR (2014). A CGG-repeat expansion mutation in ZNF713 causes FRA7A: association with autistic spectrum disorder in two families. *Human Mutation* 35, 1295–1300. [PubMed: 25196122]
- Miwa JM, Stevens TR, King SL, Caldarone BJ, Ibanez-Tallon I, Xiao C, ... Heintz N (2006). The Prototoxin lynx1 Acts on Nicotinic Acetylcholine Receptors to Balance Neuronal Activity and Survival In Vivo. *Neuron*, 51, 587–600. [PubMed: 16950157]
- Morishita H, Miwa JM, Heintz N, & Hensch TK (2010). Lynx1, a cholinergic brake, limits plasticity in adult visual cortex. *Science*, 330, 1238–1240. [PubMed: 21071629]
- Mowry BJ, & Gratten J (2013). The emerging spectrum of allelic variation in schizophrenia: current evidence and strategies for the identification and functional characterization of common and rare variants. *Molecular Psychiatry*, 18, 38–52. [PubMed: 22547114]
- O'Dushlaine C, Komiyama NH, Solovieff N, Shakir K, Fernández E, Purcell SM, ... Kähler A (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, 506, 185–190. [PubMed: 24463508]
- O'Roak BJ, Fisher SE, Baker C, Schwartz JJ, Karakoc E, Rieder MJ, ... Nickerson DA (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics*, 43, 585–589. [PubMed: 21572417]
- Person RE, Duker AL, Rosenfeld JA, Bejjani BA, Lamb AN, Bawle EV, ... Sahoo T (2010). Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *European Journal of Human Genetics*, 18, 1196–1201. [PubMed: 20588305]
- Rieusset A, Schaller F, Unmehopa U, Matarazzo V, Watrin F, Linke M, ... Muscatelli F (2013). Stochastic Loss of Silencing of the Imprinted Ndn/NDN Allele, in a Mouse Model and Humans with Prader-Willi Syndrome, Has Functional Consequences. *PLoS Genetics*, 9, e1003752. [PubMed: 24039599]
- Sanchez-Delgado M, Riccio A, Eggermann T, Maher ER, Lapunzina P, Mackay D, & Monk D (2016). Causes and Consequences of Multi-Locus Imprinting Disturbances in Humans. *Trends in Genetics*, 32, 444–455. [PubMed: 27235113]
- Sarafidou T, Kahl C, Martinez-Garay I, Mangelsdorf M, Gesk S, Baker E, Kokkinaki M, Talley P, Maltby EL, French L, Harder L, Hinzmann B, Nobile C, Richkind K, Finnis M, Deloukas P, Sutherland GR, Kutsche K, Moschonas NK, Siebert R, Gécz J, European Collaborative Consortium for the Study of ADLLE. (2004). Folate-sensitive fragile site FRA10A is due to an expansion of a CGG repeat in a novel gene, FRA10AC1, encoding a nuclear protein. *Genomics*, 84, 69–81. [PubMed: 15203205]
- Shah S, Martin NG, Painter JN, Hemani G, Bowdler L, Montgomery GW, ... Henders AK (2014). Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biology*, 15, R73. [PubMed: 24887635]
- Shohat S, Ben-David E, & Shifman S (2017). Varying Intolerance of Gene Pathways to Mutational Classes Explain Genetic Convergence across Neuropsychiatric Disorders. *Cell Reports*, 18, 2217–2227. [PubMed: 28249166]
- Sklar P, Gaugler T, Manaa D, Bodea CA, Sanders SJ, Buxbaum JD, ... Lee AB (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, 46, 881–885. [PubMed: 25038753]
- Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM, ... Sullivan PF (2014). Copy number variation in schizophrenia in Sweden. *Molecular Psychiatry*, 19, 762–773. [PubMed: 24776740]
- Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, & Beck S (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*, 29, 189–196. [PubMed: 23175756]
- Turner CLS, Aljareh S, Chi Dung V, Lokulo-Sodipe O, Mehta SG, Patalan M, ... Temple IK (2018). Maternal variants in NLRP and other maternal effect proteins are associated with multilocus imprinting disturbance in offspring. *Journal of Medical Genetics*, 55, 497–504. [PubMed: 29574422]

- Winnepenninckx B, Debacker K, Ramsay J, Smeets D, Smits A, FitzPatrick DR, Kooy RF (2007). CGG-repeat expansion in the DIP2B gene is associated with the fragile site FRA12A on chromosome 12q13.1. *American Journal of Human Genetics*, 80, 221–31. [PubMed: 17236128]
- Wormley BK, Holmans PA, Kim Y, Corvin A, Pulver AE, Shetty A, ... Kendler KS (2016). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics*, 49, 27–35. [PubMed: 27869829]
- Yakhini Z, Eden E, Navon R, Steinfeld I, & Lipson D (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, 10, 48. [PubMed: 19192299]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Clusters of *cis*-linked epivariations occur at increased frequency in cases of autism/schizophrenia compared to controls.

We identified several individuals who carried multiple epivariations clustered *in cis*, including genes with known roles in neurodevelopmental phenotypes. **(A)** Two epivariations at *GIGYF1*, a gene that shows a significant enrichment for damaging coding variants in samples with autism (J. Buxbaum, personal communication). **(B)** Two regions of hypomethylation intragenic within *SHANK3*, a gene which shows recurrent mutations in patients with autism (Boccuto et al., 2013). **(C)** Three hypermethylated CpG islands located upstream of *LYNX1*, *THEM6* and *LOC100288181* in an individual with schizophrenia. *LYNX1* is predominantly expressed in the brain and has been implicated in maintaining the stability of cortical networks and neuronal plasticity (Miwa et al., 2006; Morishita et al., 2010). In each plot, the shaded regions indicate the epivariations. Below each plot are screenshots from the UCSC Genome Browser, showing annotations for UCSC genes and CpG islands.

Figure 2. An enrichment of rare epivariations identified at imprinted loci.

We identified several individuals who showed altered methylation at imprinted loci that show parental-specific DNA methylation marks. Examples of these include: **(A)** Hypermethylation at the *SGCE/PEG10* imprinted region in an individual with schizophrenia. **(B)** Hypomethylation of both the *NDN* and *MAGEL2* imprinted regions in an individual with schizophrenia. **(C)** Hypomethylation of the *NDN* imprinted region in an individual with autism. In **(B)** the shaded regions indicate the epivariations. Below each plot are screenshots from the UCSC Genome Browser showing annotations for UCSC genes.

Figure 3. Genes associated with rare epivariations in autism cases show increased brain expression.

Utilizing data for 12 brain regions studied by the GTEx project, we compared the median expression levels of genes associated with epivariations identified in the autism cohort (dark grey bars), schizophrenia cohort (light grey bars), and the cohort of 2,711 population controls (black bars). Epivariations identified in samples with autism showed higher levels of brain expression compared to the null across all brain regions. In contrast epivariations found in population controls showed no evidence of association with high brain expression. Single asterisks indicate $p < 0.05$, double asterisks indicate $p < 0.01$. Full details are shown in Supp. Table S4.